International Journal of Innovative and Emerging Research in Engineering
Volume 2, Issue 3, 2015

Available online at www.ijiere.com

International Journal of Innovative and Emerging Research in Engineering

e-ISSN: 2394 – 3343 p-ISSN: 2394 – 5494

Mobile Visual Search: A Low Transmission Overhead Framework Based on Vocabulary Decomposition

Ms. Priti S. Khaire
Computer Science & Engineering, SSGBCOET,Bhusawal.

ABSTRACT:
In this paper, I have discuss the sub key point of the new changed framework vector quantization to move around which in carrier of illustration words model from the server to the client. In this paper in Mobile Visual Search (MVS) I see the limitation of the bandwidth in wireless networks, the big problem is transmission above your head framework. There are many kinds of compressed descriptors which are new changed framework, also manipulative a suitable lossless compacted descriptor, Has been proven elusive, In this there is no matter what descriptors are used in frame, the client transmits only the ID numbers of the illustration(visual) wording to the server, thereby reach the negligible possible transmission above your head framework.

Keywords: Bag of visual words, joint optimized product quantization, mobile visual search, vector quantization.

I. INTRODUCTION

Mobile computing is Development with an application, in which Mobile Visual Search (MVS) is very popular. MVS can be used for, shopping comparison movie poster retrieval, landmark identification, and so on. Generally, MVS is viewed as a typical example of Content Based Image Retrieval (CBIR).Most of the previous MVS systems adopt the popular framework of CBIR based on the Bag Of Visual Words (BOVW) model. In this framework, the client captures a query image and then transmits it to the server for retrieving similar images. When receiving the query image from the client, local feature descriptors like SIFT [3](D. G. Lowe) are extracted to represent the query image content. Then, these descriptors are quantized to visual words in a given vocabulary. Finally, I can search similar images which contain the same visual words as the retrieval results.

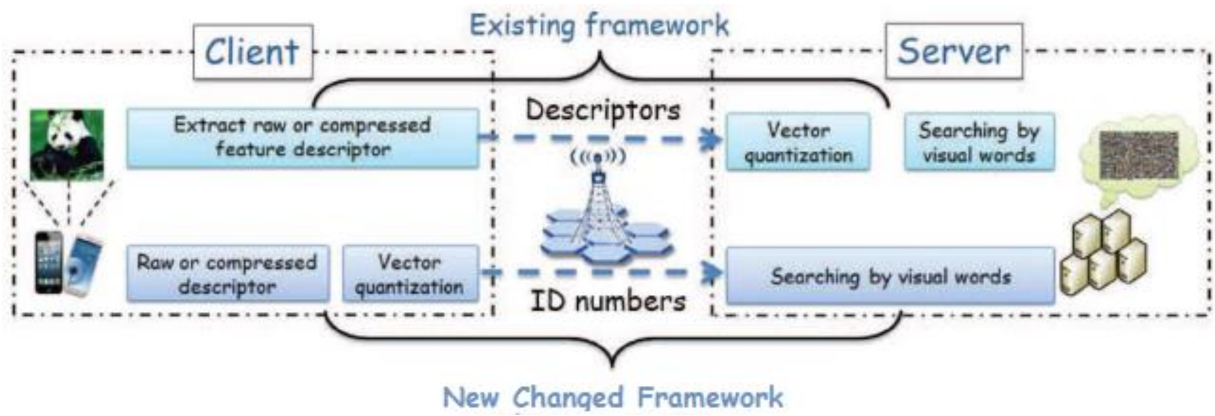

Figure 1. The differences between the existing framework and the new changed framework [1]

Although the above framework is feasible, it ignores some particular problems in MVS, especially the problem of transmission overhead. Because bandwidth is limited in wireless networks, the size of transmission data in MVS should be as small as possible to consume less bandwidth in order to reduce network latency [4](y.c.su). Therefore, feature extraction is implemented on the client. The client transmits local feature descriptors to the server instead of the query image. However, transmitting raw local feature descriptors may cost more bandwidth than transmitting the compressed image at times. It is necessary to compress local feature descriptors to reduce the transmission overhead framework. Recent work in MVS focuses on designing compressed local feature descriptors. However, existing work shows that compressed descriptors affect the retrieval performance [6](V. Chandrasekhar).

Moreover, the transmission overhead is still high when hundreds of compressed descriptors are extracted from an image. It remains to be studied how to further reduce the transmission overhead.

II. RELATED WORK

In this we can three type which is as follows:

A. The State-of-the-Art Framework of MVS

This framework consists of both an off-line part and an on-line part. The off-line part includes vocabulary building, vector quantization and inverted index construction. The on-line part includes query data transmission, vector Quantization and similar image searching. In the off-line part, vocabulary building is to acquire cluster centers by performing k-means clustering on a large number of sample descriptors extracted from images in the database. The cluster centers are viewed as the visual words in the vocabulary. Then, the vector quantization is to encode every descriptor to the nearest visual word in the vocabulary, thereby representing any one image in the database by a set of visual words. Finally, an inverted index is constructed, in which each visual word points to a list of image swhich contain it. This inverted index functions like the inverted file in textual information retrieval. In the on-line part, the client first transmiting the query data to the server. And To reduce the transmission overhead framework, the client transmits the local feature descriptors extracted from the query image as the query data. When receiving the descriptors from the client, the server encode them into visual words in the vocabulary by vector quantization as well. By searching the inverted index, images containing the same visual words from the query can be found as candidate results. By computing the similarities between the query and the candidate results, the final retrieval results are acquired. Finally, the server send the final retrieval results to the client.

B. Feature Descriptor Compression

In the state-of-the-art framework of MVS, to reduce transmission overhead, the client transmits local feature descriptors extracted from the query image to the server instead of transmitting the whole query image. However, transmitting some raw local feature descriptors may cost more bandwidth than transmitting the compressed image. When one image is represented by a set of SIFT descriptors, the size of the descriptors is typically larger than the size of the compressed image. To address this problem, descriptor compression methods are of increasing interest in the area of MVS. The existing descriptor compression methods fall into two categories. The first compresses descriptors by dimensionality reduction or feature coding, such as the method based on Principle Component Analysis (PCA) [8](Y. Ke & R. Suthankar), the method based on Linear Discriminant Analysis (LDA) [9](S.W.G. hua & M. Brown), hash coding [10](Y. Wejss), and so on. The second designs the compressed descriptors directly.

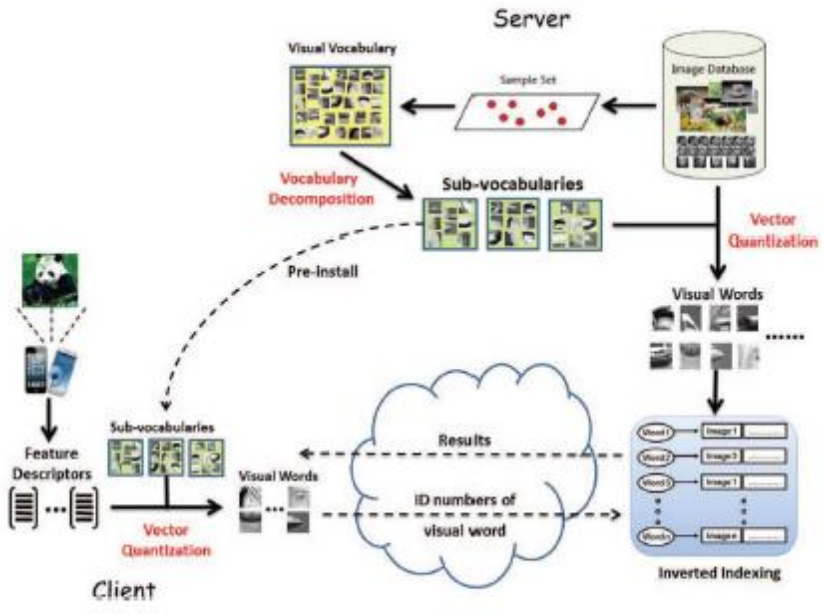

Figure 2.The framework of MVS based on vocabulary decomposition [1]

Existing work shows that simply compressing the descriptors affects the performance of image retrieval, because dimensionality reduction and feature coding lead to information loss[12](V.chandrasekhar ,G.Takacs), [13](D.M.Chen). It is difficult to design highly discriminative compressed descriptors without information loss. Existing work shows that the CHoG descriptor is a better choice in MVS, which outperforms other descriptors when transmitting lower or equivalent bit rates. However, hundreds of CHoG descriptors are usually extracted from an image with complex content. The transmission overhead is still high. How to further reduce the transmission overhead is a problem.

C. Vector Quantization

Vector quantization is the core of the framework based on the bag of visual words model. It is the basis of inverted index construction and similar image searching. In vector quantization, the descriptors are viewed as similar if they are encoded into the same visual word, and dissimilar otherwise. Thus vector quantization provides a very coarse approximation to the actual distances between descriptors, which leads to information loss [14](J.Philbin, O.Chum). To reduce the information loss in the vector quantization, a large vocabulary is necessary. However, traditional k-means clustering is difficult to scale to a large vocabulary. A large vocabulary is constructed by computing the Cartesian Product of all sub-vocabularies. When encoding one descriptor to a visual word in the large vocabulary, the descriptor is first encoded into multiple sub-words in sub-vocabularies. Then the combination of these sub-words is mapped to a visual word in the large vocabulary. To further reduce the distances between the original descriptors and the visual words in the large vocabulary (the information loss), Optimized Product Quantization (OPQ) is new changed [7](T. Ge).

III. THE VOCABULARY DECOMPOSITION-BASED FRAMEWORK OF MVS

In the existing framework based on the bag of visual words model, the transmission overhead is still a problem. If I can migrate the vector quantization from the server to the client, the client only needs to transmit the ID numbers of visual words to the server. In this case the transmission overhead is further reduced without designing new compressed descriptors. However, due to the limited memory on mobile devices, a large vocabulary for vector quantization cannot be stored on the client. To address this problem, we new change the vocabulary decomposition and a new framework of MVS based on it.

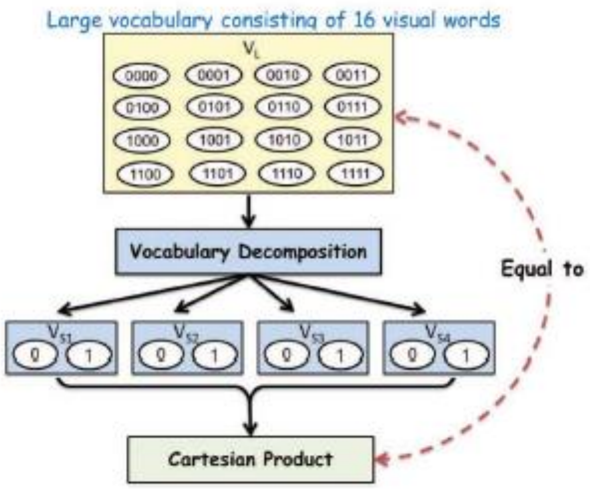

Figure 3. An example of vocabulary decomposition [1]

IV. EXPERIMENTS AND PERFORMANCE EVALUATION

A. Experimental Setup

The experiments involve comparing the performances of the state-of-the-art framework based on the bag of visual words model (BOVW framework) and the new changed framework based on vocabulary decomposition (VD framework). For objectivity of evaluation, I conduct simulation experiments and real experiments. The simulation experiments are to evaluate the VD framework based on a public image data set. The real experiments are to test the VD framework in our MVS system. In simulation experiments, I select the Stanford mobile visual search data set [6]. Compared to other popular data sets in computer vision, this data set is specially built for MVS. In this data set, there are eight different categories of images including book covers, business cards, cd covers, dvd covers and marks, museum paintings, print and video frames. Each category consists of the reference database images and the query images. The reference database images are high quality JPEG compressed color images, which are obtained from websites, key frames of video, high quality upright scans and so on. The query images are captured by different mobile devices, such as the iPhone 4, Palm, Nokia phone, Canon camera, and so on.

B. Comparison of Different Frameworks

Firstly, we evaluate the average retrieval accuracy for the BOVW framework with different sizes of vocabularies and the VD framework with different sizes of sub-vocabularies. In the VD framework, I use the Product Quantization (PQ) based method and the Joint Product Quantization (JPQ) based method for vocabulary decomposition, respectively. while the average retrieval accuracy is 84.6% for the VD framework with 4 MB sub-vocabularies (the Joint Product Quantization (JPQ) based method is used for vocabulary decomposition). Thus it is feasible to store sub-vocabularies on the client for vector quantization. Furthermore, to compare the PQ and JPQ based methods for vocabulary decomposition, we use them in the VD framework, respectively. From Fig. I can see that the VD framework with JPQ performs better than the VD framework with PQ. It illustrates that the JPQ based method is more suitable for vocabulary decomposition. Secondly, we compare the average retrieval accuracy and the transmission overhead for the BOVW framework and the VD framework. In the BOVW framework, we use SIFT descriptors and CHoG descriptors for image representation, respectively. In the VD framework, we use SIFT descriptors for image representation and JPQ for vocabulary decomposition.

C. The Effect of Parameters on Retrieval Performance

In vocabulary decomposition, a large vocabulary is decomposed into sub-vocabularies. From Section III-C, I can see that the number of entries in the inverted index is directly proportional to the value of *m*. If there are more entries in the inverted index, the chance of different images appearing in the same entry can be reduced. However, if the value of is large, the dimension of the sub-word is small resulting in poor discrimination. Thus the value of affects the performance of retrieval. To test the effect of the value of *m* we decompose a vocabulary consisting of visual words into 2, 4, and 8 sub-vocabularies respectively. Fig. shows the average retrieval accuracies of each category in the data set for the VD framework with different numbers of sub-vocabularies, in which the PQ based method is used for vocabulary decomposition. The average retrieval accuracy is the best when using 2 sub-vocabularies. It illustrates that it is not preferable to decompose the vocabulary into many sub-vocabularies. Although a large inverted index can be built based on many sub-vocabularies to reduce the chance of different images in the same entry, short sub-words affect the performance more obviously.

D. The Comparison of Vocabulary Decomposition Methods

I compare different vocabulary decomposition methods here. Based on existing work, I have the PQ, OPQ-P and OPQ-NP based methods. To reduce information loss in them, I new change the JPQ, JOPQ-P and JOPQ-NP based methods. I use these six methods in the VD framework for vocabulary decomposition, respectively. In this experiment, I decompose a vocabulary consisting of visual words into 2 sub-vocabularies and each consisting of visual words.

E. The Real Experiments on MVS System

To further evaluate the new changed VD framework, I develop an MVS system for movie poster retrieval. In this system, I first capture a movie poster with a mobile device. Then, I send query data to the server in order to find information about the movie. Fig. shows an example of our MVS system. This MVS system consists of the client and the server. The client is an MI 2S mobile phone with a quad-core 1.7 GHz CPU, 2G RAM, 8 Megapixel camera and Android OS. The server is running on a computer with Intel(R) Xeon(R) dual processor (2.4 GHz), 12 GB memory, 25 TB disk, and 64-bit Windows OS. First, I measure the transmission time of different frameworks in a WIFI network. I randomly select ten pictures from our movie poster database as the query images. In the BOVW framework, Then I transmit SIFT descriptors. In the VD framework, I transmit the ID of visual words. From the measured results, I can see that the VD framework can reduce the transmission time significantly. For some pictures, the VD framework outperforms the BOVW framework by reducing more than 95% of the transmission time.

CONCLUSIONS

In this framework, I want to implement the vector quantization on the client to reduce the transmission overhead. In this paper, I new change a new framework based on vocabulary decomposition for MVS. To achieve this goal, I decompose a large vocabulary into small sub-vocabularies which can be stored on the client easily. To compare the new changed framework and the existing framework, I conduct simulation experiments on the Stanford MVS data set and real experiments on an MVS system. The experimental results show that the new changed framework outperforms the existing framework by reducing more than 95% of the transmission overhead. This process is called the vocabulary decomposition. I first formulate it as an optimization problem. Then I present some effective algorithms including JPQ and JOPQ for this problem. Using the new changed algorithms, I can realize effective vocabulary decomposition.

ACKNOWLEDGMENT

I feel great pleasure in submitting this review paper on "Mobile Visual Search: A Low Transmission Overhead Framework Based on Vocabulary Decomposition". I wish to express true sense of gratitude towards my teacher Mr. R. P. Chaudhari who at very discrete step in study of this paper contributed his valuable guidance and help me to solve every problem that arose. I would wish to thank our H.O.D., Prof. D. D. Patil for opening the doors of the knowledge towards the realization of the paper. Most likely I would like to express my sincere gratitude towards my family for always being there when I needed them the most. With all gratitude and respect, I would like to thank all the people, who helped me directly or indirectly. I owe my all success to them, and special thanks to all writers by whose this paper is referred by me.

REFERENCES

[1] IEEE TRANSACTIONS ON MULTIMEDIA, VOL. 16, NO. 7, NOVEMBER 2014, Heng Qi, Milos Stojmenovic, Keqiu Li, *Senior Member, IEEE*, Zhiyang Li, and Wenyu Qu "A Low Transmission Overhead Framework of Mobile Visual Search Based on Vocabulary Decomposition"

[2] In *9th IEEE Int. Conf. Computer Vision (ICCV)*, 2003, pp. 1470–1477. J. Sivic and A. Zisserman, "Video Google: A text retrieval approach to object matching in videos,".

[3] *Int. J. Comput. Vision (IJCV)*, vol. 60, no. 2, pp. 91–110, 2004, D. G. Lowe, "Distinctive image features from scale-invariant keypoints," .

[4] In *Proc. 21st ACM Int. Conf. Multimedia (MM)*, 2013, pp. 73–82, Y. C. Su *et al.*, "Enabling low bitratemobile visual recognition—Aperformance versus bandwidth evaluation".

[5] In *Proc. IEEE Int. Conf. Computer Vision and Pattern Recogn. (CVPR)*, 2009, pp. 1–8. V. Chandrasekhar *et al.*, "CHoG: Compressed histogram of gradients— A low bit rate feature descriptor

[6] In *Proc. 2nd ACM Conf. Multimedia Systems (MMSys)*, 2011, pp.117–122,V. Chandrasekhar *et al.*, "The stanford mobile visual search data set," .

[7] In *Proc. IEEE Int. Conf. Computer Vision and Pattern Recogn. (CVPR)*, 2013, pp. 1–8. T. Ge *et al.*, "Optimized product quantization for approximate nearest neighbor search,"

[8] In *Proc. IEEE Int. Conf. Computer Vision and Pattern Recogn. (CVPR)*, 2004, pp. 506–513. Y. Ke and R. Sukthankar, "PCA-SIFT: A more distinctive representation for local image descriptor," .

[9] In *Proc. 11th IEEE Int. Conf. Computer Vision (ICCV)*, 2007, pp. 1–8. S. W. G. Hua and M. Brown, "Discriminant embedding for local image descriptors,"

[10] In *Proc. Advances in Neural Information Processing Systems (NIPS)*, 2008, pp. 1–8,Y. Weiss, A. Torralba, and R. Fergus, "Spectral hashing,"

[11] *Comput. Vision Image Understand.*, vol. 110, no. 3, pp. 346–359, 2008. H. Bay, A. Ess, T. Tuytelaars, and L. V. Gool, "Speeded-up robust feature,"

[12] In *Proc. Int. Mobile Multimedia Workshop (IMMW) in conjunction with IEEE Int. Conf. Pattern Recogn. (ICPR)*, 2010. V. Chandrasekhar, M. Makar, G. Takacs, D. M. Chen, S. S. Tsai, N. M. Cheung, R. Grzeszczuk, Y. Reznik, and B. Girod, "Survey of SIFT compression schemes,"

[13] In *Proc. IEEE 17th Int. Conf. Image Process. (ICIP)*, 2010, pp. 3885–3888. V. Chandrasekhar, D. M. Chen, A. Lin, G. Takacs, S. S. Tsai, N. M. Cheung, Y. Reznik, R.Grzeszczuk, and B.Girod, "Comparison of local feature descriptors for mobile visual search," .

[14] In *Proc. IEEE Int. Conf. Computer Vision and Pattern Recogn. (CVPR)*, 2008, pp. 1–8,J. Philbin, O. Chum, M. Isard, J. Sivic, and A. Zisserman, "Lost in quantization: Improving particular object retrieval in large scale image databases,".